

University of Groningen

Modelling for markets. Applications of advanced models and methods for data analysis

Plat, Francis Wilhelmus

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

1988

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Plat, F. W. (1988). *Modelling for markets. Applications of advanced models and methods for data analysis*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CHAPTER 8

SUMMARY AND POSTSCRIPT

8.1 Introduction

This study deals with some aspects of the specification and estimation of marketing models. Having witnessed three stages in the theory and practice of marketing model building, the models discussed in this study may be considered to refer to a fourth era in marketing model building. The specification of these models arose from a desire to develop models that account for factors like:

- the presence of errors in the data which are used,
- the effects of actions and reactions of competitors, and
- possible structural shifts in consumer response over time.

The study can be broadly divided into two parts. The first part deals with data. The sources of errors or bias in audit data as well as the consequences of these errors are examined. Remedial actions are discussed. One of these remedies involves the specification of models that account for possible errors in the data. Finally it is examined whether or not things may change for the better if scanners are used for data collection.

The second part of this study relates to the development of models that account for specific market phenomena. Several extensions of models which account for competitive actions and reactions are examined. Furthermore, models are used to assess possible structural shifts in consumer response over time. Doing so it is analyzed whether or not models need to be adapted to cover circumstances which are present in an era of stagflation. In what follows the data and the models part of this study are summarized.

8.2 Data

Audit data such as consumer panel data and retail store audit data constitute an important set of secondary external data. These data, however, may be subject to errors or bias. The sources and the consequences of the data bias in audit data have been examined in chapter 2.

Four sources of data bias in audit data can be distinguished, i.e. statistical bias, sample bias, non-response bias and survey bias. For the data which are used in this study only the non-response bias and a particular form of survey bias appeared to be relevant. With respect to the consumer panel it was found that consumers buying relatively cheap brands have a higher response rate than those buying the more expensive brands. As far as the store audit is concerned the non-response bias results from the overproportional refusal of discounters to participate in the audit. For the store audit a form of survey bias, referred to as instrument bias, is also relevant. The instrument bias stems from different stores being audited on different days, in different weeks, or even in different months. Clearly, the instrument bias in store audit data is particularly relevant when short-term, e.g. monthly or bi-monthly, fluctuations are concerned.

As a consequence of the data bias considerable differences may occur between consumer panel data and store audit data, which in turn may lead to different model outcomes. The differences between consumer panel data and store audit data have been illustrated empirically using data of market 1 to determine so-called coverage factors. Market 1 represents the Dutch market of a frequently purchased low-priced non-durable consumer good. In these factors audit data on sales or market shares are related to corresponding internal data. The non-response bias in consumer panel data is reflected in an apparent inverse relation between consumer panel coverage factors and prices. For the store audit no relation between coverage and marketing variables was found.

By numerically specifying market share models it was investigated to what extent the use of data from different audits leads to different descriptions of the relations between marketing response measures and marketing instruments. In marketing practice it has been argued that variables measured by different audits may differ in average value but will show similar fluctuations. No strong evidence in favor of this argument was found. For five of the twelve brands we have analyzed, the estimates of the response parameters, obtained using data from different audits, were significantly different.

Finally, methods to reduce the data bias in consumer panel- and store audit data have been discussed. The data bias in consumer panel data on (brand) sales and market shares can be reduced by one third using a formal analysis in which the coverage factors of brand sales or market shares are related to the relative prices and advertising shares

of the brand
reached by

The ana
ce of the d
the paramet
inconsisten
3 and 4 ain
used, thus
resulted in
respectivel
first two m
variables o
measurement

All thr
ces which p
audit data.
models. A g
given in cl

Startin
model pres

$$m_{jt} = \beta_{0j} -$$

$$u_{jt},$$

where:

$$m_{jt} = \text{mar}$$

bran

$$rp_{jt} = \text{rel}$$

div

(se

$$d_{jt} = \text{eff}$$

$$ap_{jt} = \text{adv}$$

$$ar_{jt} = \text{adv}$$

$$at_{jt} = \text{adv}$$

$$u_{jt} = \text{ran}$$

$$n = \text{num}$$

$$\beta's = \text{unk}$$

of the brands. A reduction of the data bias in store audit data can be reached by applying a reweighing procedure.

The analyses performed in chapter 2 illustrate that as a consequence of the data bias, the use of audit data may cause the estimators of the parameters of a marketing model as obtained by OLS to be biased and inconsistent. Instead of reducing the data bias in audit data, chapters 3 and 4 aim at specifying models that account for errors in the data used, thus improving the quality of the parameter estimates. This has resulted in a trilogy of approaches to the problem of data bias, namely respectively a limited-, extended- and full information approach. The first two methods account for measurement errors in the explanatory variables only, whereas the third method also explicitly deals with measurement errors in the dependent variable.

All three methods yield estimates of the measurement error variances which provide us with information regarding the reliability of the audit data. The models involved are Linear Structural RELation (LISREL) models. A general description of the specification of LISREL-models is given in chapter 4.

Starting point of the analysis is the linear additive market share model presented in chapter 2 as relation (2.6):

$$m_{jt} = \beta_{0j} + \beta_{1j}rp_{jt} + \beta_{2j}d_{jt} + \beta_{3j}\frac{ap_{jt}}{\sum_{i=1}^n ap_{it}} + \beta_{4j}\frac{ar_{jt}}{\sum_{i=1}^n ar_{it}} + \beta_{5j}\frac{at_{jt}}{\sum_{i=1}^n at_{it}} + u_{jt}, \quad (8.1)$$

where:

m_{jt} = market share (defined with respect to the market or segment) of brand j in period t ,

rp_{jt} = relative price of brand j in period t , i.e. the price of brand j divided by the weighed average price of all brands on the market (segment), using the brand sales as weighing factors,

d_{jt} = effective store distribution (fraction) of brand j in period t ,

ap_{jt} = advertising expenditures on press for brand j in period t ,

ar_{jt} = advertising expenditures on radio for brand j in period t ,

at_{jt} = advertising expenditures on television for brand j in period t ,

u_{jt} = random disturbance term,

n = number of brands on the market (segment),

β 's = unknown parameters.

In this model the market share of a brand, as measured by a consumer panel, is related to the relative price of the brand (as measured by a consumer panel), the effective store distribution of the brand (as measured by a store audit) and a number of advertising variables (as measured by an advertising audience agency). It is assumed that the price- and distribution variables are measured with error and that the advertising variables are measured correctly.

In the limited information approach consistent estimators of the parameters of the market share model have been determined by modifying relation (8.1) to represent a linear errors-in-variables model. That is the observed values of the erroneously measured explanatory variables are expressed in terms of the true values of the explanatory variables and measurement errors.

Assuming the measurement errors in the observed values of the explanatory variables to be uncorrelated and using alternative measurements of the erroneously measured explanatory variables as 'instruments' or 'instrumental variables', the model becomes identifiable. In this study the relative price of the brand as measured by a store audit and the market coverage of the brand as measured by a store audit have been used as instrumental variables corresponding with the erroneously measured explanatory variables mentioned above.

The values of the resultant estimators have been referred to as 'instrumental variable estimates' or 'IV-estimates'. Consistent estimates can be obtained of:

1. the response parameters,
2. the variance of the disturbance term,
3. the variances of the measurement errors in the observed values of the explanatory variables, and
4. the variances and covariances of the true values of the explanatory variables.

The relative merits of various ways of estimating the standard errors of the estimators of the parameters 1 - 4 above have been discussed. In particular the jackknife and the bootstrap have been applied. Both methods have also been used to determine whether or not the IV-estimates, though consistent, are biased. The latter analysis revealed that a number of IV-estimates is substantially biased and needs to be corrected for bias. Compared with the significance tests based on the bias-corrected estimates it appeared that, grosso modo, the use of the uncorrected IV-estimates does not give misleading results.

For some
ferred subst
and the boo
differences

The abo
in relative
biased. As
parameters
balance of t
below. This
fication of
tions betwe
to be suffi
explanatory

The est
without pro
cally speci
analyzed be
sidered too
and distrib
significant
appointing
were found
It appeared
that for the
natory vari
five brands
errors in t
correlation
variables m
show any of
results obt

From th
OLS- and IV
to some ext
ror varianc
noise-to-si
defined as
population

For some brands the IV-estimates of the response parameters differed substantially from the corresponding OLS-estimates. The jackknife and the bootstrap, however, revealed that, except for two brands, the differences are not statistically significant.

The above can, in short, be stated as follows: as can be expected in relatively small samples, consistent estimators can be substantially biased. As a consequence, the OLS- and IV-estimates of the response parameters can be quite similar. A second explanation for the resemblance of the OLS- and IV-estimates of the response parameters is given below. This explanation is based on the fact that the numerical specification of a linear errors-in-variables model requires the correlations between the explanatory variables and the instrumental variables to be sufficient. In this study this means that the reliability of the explanatory variables must be sufficient.

The estimation of the linear errors-in-variables model was not without problems. Linear errors-in-variables models have been numerically specified for nine brands on market 1. Three brands have not been analyzed because either the number of observations available was considered too small or because, in chapter 2, the estimates of the price- and distribution coefficients in relation (2.6) were not statistically significant. For five of the nine brands the results were rather disappointing in the sense that unduly large estimates of standard errors were found or that unduly large negative variance estimates resulted. It appeared that these findings can partly be explained from the fact that for these five brands some of the correlations between the explanatory variables and instrumental variables are relatively low. For five brands, therefore, the model was modified ignoring the measurement errors in the explanatory variables corresponding with relatively low correlations. Doing so, for two of these five brands an errors-in-variables model was obtained of which the estimation results did not show any of the deficiencies mentioned above. Thus, the estimation results obtained for six brands remained to be analyzed.

From the estimation results it appeared that the resemblance of the OLS- and IV-estimates of the response parameters mentioned above can, to some extent, be explained from the estimates of the measurement error variances. From these estimates we have determined estimates of the noise-to-signal ratios for the explanatory variables. These ratios are defined as the quotients of the measurement error variances and the population variances of the observed values of the explanatory varia-

bles. The noise-to-signal ratios represent the fractions of the observed variances of the explanatory variables which are due to measurement errors.

It appeared that the largest differences between the OLS- and IV-estimates of the response parameters were found for those brands and explanatory variables for which the largest estimates of the noise-to-signal ratios were obtained. Furthermore larger (smaller) correlations between explanatory variables and instrumental variables appeared to correspond with smaller (larger) estimates of noise-to-signal ratios.

It is intuitively appealing to consider the correlations between the observed values of the explanatory variables and the corresponding alternative measurements as indications for the degree in which the true explanatory variables can be reliably measured. Thus, the larger these correlations are, the more reliable the explanatory variables are measured and hence, the smaller the estimates of the noise-to-signal ratios and the differences between the OLS- and IV-estimates of the response parameters will be. The results also show that there is a threshold below which the correlations are too small to successfully estimate the parameters of a linear errors-in-variables model¹.

In the linear errors-in-variables model discussed so far, the instrumental variables are related to all explanatory variables in the model and not just to the explanatory variables they measure. In other words no a priori information regarding the relationships between instrumental variables and explanatory variables is taken into account. For this reason and since alternative measurements of the dependent variable are not incorporated in the model the numerical specification of models of this kind has been referred to as a limited information approach to the problem of data bias.

In the extended information approach the relationships between instrumental variables and explanatory variables are modified so as to account for the a priori information regarding the relationships between the two categories of variables. Besides the reliability of the observed values of the explanatory variables now also the reliability of the instrumental variables can be determined. Since in this study the instrumental variables and observed explanatory variables both are measurements of the same unobservable- or latent explanatory variables, it can now be determined which variables most reliably measure the latent explanatory variables. It appears that the ranking of measure-

ments according to means that by the consistency of other brands is the more

In the numerical simulation measurement models. Further information and unobserved

Intuitively appealing to data. From determined

The model (LISREL) named 'real' availability indicated calculated the consumer theory variable consumer perception and measurement

LISREL market level. This may indicate that the model is required

Comparing estimates in this field to mine whether information

In large the extended the observed

ments according to their reliability varies across the brands, this means that for some brands for example the relative price as measured by the consumer panel is the more reliable price measure, whereas for other brands the relative price as measured by the retail store audit is the more reliable price measure.

In the full information approach to the problem of data bias the numerical specification of the models involves the use of alternative measurements of both the dependent and the explanatory variables in the models. Furthermore, as in the extended information approach, a priori information regarding the relationships between observable variables and unobservable or latent variables is taken into account.

Intuitively this approach to the problem of data bias seems quite appealing since the parameter estimates are based on all available data. From the estimation results the reliability of the data can be determined.

The models are generally known as LInear Structural RELation (LISREL) models. The models specified in this study have been referred to as LISREL market share models. They contain three latent variables named 'real market share', 'consumer's price perception' and 'brand availability'². The latent dependent variable 'real market share' is indicated by three observable variables, namely the market share as calculated from ex-factory data and the market shares as measured by the consumer panel and the retail store audit. The two latent explanatory variables are indicated by the relative prices as measured by the consumer panel and the store audit, and the effective store distribution and market coverage as measured by the store audit respectively.

LISREL market share models have been estimated for six brands in market 1. For three brands negative variance estimates were found. This may indicate that for these brands the model is not correct or that the number of observations is too small³. In any case some caution is required regarding the interpretation of the estimates.

Compared with earlier results occasionally substantially different estimates of the response parameters were obtained. Future research in this field may involve using the jackknife or the bootstrap to determine whether or not the estimates obtained in the extended- and full information approach, though consistent, are biased.

In large, the results confirm our earlier conclusions, drawn from the extended information approach, with respect to the reliability of the observable variables corresponding with the latent explanatory

variables. Again, for different brands different rankings were found.

Although due to the occurrence of negative variance estimates the reliability of the three market share variables could not be evaluated for all brands, the market share as calculated from the ex-factory data appeared to be the more unreliable measure. This result is not surprising since, as has been pointed out in chapter 2, particularly if small data intervals are used, the corrected ex-factory sales may be quite unreliable⁴.

Having discussed and analyzed the reliability of the traditional audit data, in chapter 5 we have addressed some recent developments in marketing research. More specifically, the development of scanning and its implications for marketing and marketing research have been discussed. Special attention has been given to the question whether or not by using scanners the problem of data bias in audit data can be solved.

Scanning offers numerous benefits to consumers and retailers. The development of systems for automated transmission of transaction data may lead to additional savings for industry and trade. With scanning an important new measurement instrument has emerged. Describing the scanning-based services being currently operational in the U.S., a picture is given of some of the opportunities that scanning, and the combination of scanning with other technologies, offer for marketing research.

Scanning-based audits have been compared with the traditional retail store audits and consumer panels. By scanning consumer identity cards or by having panel members scan their purchases at home, scanning-based consumer panels can be constructed. The main benefits of scanning-based store audits are: greater accuracy, lower costs of data collection, shorter data intervals, no instrument bias, increased speed of reporting and more data. The main benefits of scanning-based consumer panels are: greater accuracy, increased speed of reporting and more data.

With respect to the problem of data bias in audit data it was found that one of the benefits of using scanners is that the instrument bias in store audit data is eliminated. Dependent on the retailer's willingness to participate in the audit, the non-response bias, however, may still be substantial. As far as the consumer panels are concerned the use of scanners implies a reduction of the work load of the respondents. As a consequence response rates may increase and, hence, the non-response bias in the data may reduce. As far as scanning-based consumer panels involve retailers scanning identity cards and providing

the da
be sub
data.

Wi
abilit
mathem
have b
observ
numeric
market
pendent
adverti
ticipat

The
needed,
sing an
data sh
store X
retailer
ferent

8.3 Mod

In
dels ha
chapter
account

In
for the
of the
veloped
based o
In this
leader
tions o
models
between

the data, the traditional non-response bias in consumer panel data may be substituted by the non-response bias which occurs in store audit data.

With the growing penetration of scanners and the increasing availability of data, retail management may become more inclined to use mathematical models for decision support. Thusfar relatively few models have been developed from the retailer's point of view. Using weekly observations of a number of brands on market 1 such a model has been numerically specified. The dependent variable in the model is the market share of a brand as measured in one store (store X). The independent variables are: the relative price (as measured in store X), the advertisements made by store X and by competitive stores, and the participation of store X in consumer promotions.

The estimation results indicated that data on more variables are needed, such as data on shelf space allocation, stock-outs, merchandising and different types of consumer promotions. Furthermore, these data should also be available of the competitive stores and not only of store X. Put otherwise, the determination of the effectiveness of the retailer's marketing instruments requires a pooling of data of different stores.

8.3 Models

In the data part of this study the specification of marketing models has been centered around the problem of data bias in the data. In chapters 6 and 7 we have considered the specification of models that account for specific market phenomena.

In chapter 6 we have examined how models can be extended to account for the effects of actions and reactions of competitors. Starting point of the analysis is the normative model of competitive behavior developed by Lambin, Naert and Bultez (1975): LNB-model. This model is based on the leader-follower reaction theory of oligopolistic markets. In this model the reactions of the followers on the actions of the leader are not specified for each individual brand, rather the reactions of competition as a whole are considered. In the generalized LNB-models which have been developed in chapter 6 a distinction is made between the reactions of the competitors in different segments on

actions of the leader. This means a decomposition of the competitive reactions.

Versions of the LNB-model and of the generalized LNB-models have been estimated using bi-monthly observations of market 1. In this oligopolistic market three segments are distinguished. It is assumed that the leader operates on one segment only. The parameterization of the models shows that the direct effects of changes in the marketing instruments on the leader's sales are partly compensated and partly reinforced by the indirect effects of the reactions of the followers in the leader's segment. The generalized LNB-models of competitive marketing behavior incorporate a decomposition of the indirect effects of the reactions of competitors in the segments on which the leader does not operate. In our study the reactions of these competitors do not effect the sales elasticities of the leader. The sales elasticities are determined to an important degree by the direct effects the marketing instruments have on market share of segment sales. The findings illustrate that, as compared with the LNB-model, the numerical specification of the generalized models may offer additional insights into the effects the marketing instruments of the leader ultimately have on brand sales.

These insights could be further increased by applying estimation methods such as ridge regression or latent root regression, to account for multicollinearity.

In chapter 7 we have investigated whether or not marketing models need to be adapted to account for circumstances which are present in an era of stagflation. The hypotheses have been formulated that during a period of stagflation the price- and distribution elasticities will increase in absolute value and that the (theme) advertising elasticities will decrease. We have investigated whether or not these changes are reflected in changes of the parameter estimates, or can be measured otherwise. Market share models have been estimated for 19 brands on five different markets.

The effects of stagflation on consumer response have been measured in four different ways. First, changes in the elasticities were measured using relatively simple models in performing piecewise regressions and moving window regressions. Then, an environmental variable approach was applied in which the effects of stagflation on consumer response are measured by estimating consumption elasticities. Finally,

more co
been ar
Bot
indicat
change
methods
the pat
as expe
terns w
of incr
(theme)
of bran
small n
signifi
Thu
moving
tively
number
ted. As
the obs
respons
The
ly hamp
consump
figures
of the
found t
Negativ
found f
cheap.
expensi
The
worthwh
approac
that ac
Col
ted the
brands,
were th

more complex models which contain consumption varying elasticities have been analyzed.

Both the piecewise regressions and the moving window regressions indicated that the estimates of the elasticities may substantially change over time. The different estimates as obtained by the two methods in general followed similar patterns. For a number of brands the patterns followed by one or more of the estimated elasticities were as expected. However, for about an equal number of brands opposite patterns were found. Covariance analysis revealed that the hypotheses of increasing price- and distribution elasticities and decreasing (theme) advertising elasticities were accepted for only a small number of brands. For most brands the hypotheses were rejected, whereas for a small number of brands opposite patterns were found to be statistically significant.

Thus it may be concluded from the piecewise regressions and the moving window regressions that the estimates of the parameters of relatively simple models can easily change over time. Only for a minor number of brands, however, changes have been found that were as expected. Assuming our hypotheses to hold, the question rises whether or not the observed changes really reflect structural shifts in consumer response.

The application of the environmental variable approach was seriously hampered by the extreme collinearity between price variables and consumption figures and between distribution variables and consumption figures. This collinearity considerably complicated the determination of the consumption elasticities. For 7 brands the market shares were found to be significantly effected by the trend in consumption figures. Negative effects of stagflation on a brand's market share have been found for brands which are either relatively expensive or relatively cheap. Positive effects were found for the cheapest brand of the most expensive product.

These results to some extent being quite appealing, it seems worthwhile to have future research conducted to further elaborate this approach. In this respect one may think of applying estimation methods that account for collinearity between explanatory variables.

Collinearity between explanatory variables also seriously complicated the estimation of the consumption varying elasticities. Only for 6 brands, for which the correlations between the explanatory variables were the lowest, the estimation results have been presented. These

results did not all confirm the results obtained by the piecewise regressions and the moving window regressions. Further research may reveal whether these differences are due to multicollinearity or not.

NOTES

Notes to c

- 1 One of
by Kei
- 2 A crit
e.g.,
stances
concep
butes
- 3 The or
(1972)
(1980)
compet
of Day
- 4 See for
specif
data as
- 5 Little
mation
tion sy
system
- 6 The abb
form.
- 7 This me
For exa
ly, a c
decisio
- 8 See als
- 9 See, e.
- 10 See, e.
- 11 See Nae
- 12 See, e.
- 13 See, e.
13.
- 14 Besides
the tru
error v
- 15 In fact
3 const